Proceedings of the
Indian Statistical Institute Golden Jubilee International Conference on
Statistics : Applications and New Directions
Calcutta 16 *December*—19 *December*, 1981 pp. 205–236.

PARTIAL EXCHANGEABILITY AND SUFFICIENCY

By P. DIACONIS[1]
Stanford University
and
D. FREEDMAN[2]
University of California, Berkeley

SUMMARY. This paper is largely expository. A framework is presented for characterizing mixtures of processes in terms of their symmetry properties and sufficient statistics. As an application, mixtures of the following kinds of processes are characterized : coin-tossing processes (de Finetti) ; sequences of independent identically distributed normals ; sequences of independent, identically distributed integer-valued generalized exponential variables.

1. Introduction

Let $X_1, X_2, \ldots,$ be a sequence of zero-one valued random variables. They are *exchangeable* if their joint distribution is invariant under finite permutations of the indices. An equivalent formulation : let $S_n = X_1 + \ldots + X_n$; given $S_n = t$, conditionally the sequence $X_1, \ldots, X_n$ is uniformly distributed over the $\binom{n}{t}$ sequences having t one's and $n-t$ zero's. So the X-process is exchangeable if the partial sums are sufficient, with the specified conditional distribution for $X_1, \ldots, X_n$ given S_n.

A famous theorem of de Finetti's (1931; also see 1937 or 1972) shows that $X_1, X_2, \ldots,$ are exchangeable iff this process is a mixture of coin-tossing processes, that is, for all n, and all strings $x_1, \ldots, x_n$ of 0's and 1's.

$$P(X_1 = x_1, \ldots, X_n = x_n) = \int p^t(1-p)^{n-t}\mu(dp)$$

for $t = x_1 + \ldots + x_n$. Here, μ is a probability on [0, 1], uniquely determined by P.

This theorem has been generalized in several directions. One is to allow more complex state spaces : on this score, see Hewitt-Savage (1955), Diaconis-Freedman (1980), Dubins-Freedman (1979). Another is allow more complex

[1]Research partially supported by NSF Grant MCS80-24649.
[2]Research partially supported by NSF Grant MCS 80-02535. This author also wishes to thank the University of Kuwait for its hospitality.

notions of symmetry; these are termed, collectively, "partially exchangeability". See de Finetti (1938; 1972, Sec. 9.6.2), Freedman (1962a, b), Diaconis-Freedman (1980b). Precise definitions are given below. As the discussion of the coin-tossing example hints, the statistical idea of sufficiency is relevant.

The connection between sufficiency and partial exchangeability has been explored by the Scandinavian school : see, for example, Martin-Löf (1970) and (1974), or Lauritzen (1982). The same ideas are relevant in statistical-mechanical studies of "Gibbs states" : see Lanford and Ruelle (1969), Ruelle (1978), Preston (1976) or Georgii (1979).

The key mathematical techniques involve martingales and the machinery of regular conditional distributions. Most of the technique appears in early papers by Oxtoby (1952) on the Kryloff-Bogoliouboff theory, or Hunt (1960) on the Martin boundary for Markov chains. Recently, these ideas have been put into abstract and systematic form by Lauritzen (1976) which makes the connection with the theory of the Martin boundary, and Dynkin (1978). The object of the present paper is to review the theory in a statistical context (Sections 1 and 2), and present some examples (Section 3). The material on discrete exponential families in Section 3 and some of the details in Section 4 may be new. However, this paper is largely, expository.

In the balance of this section, a general theorem on sufficiency and partial exchangeability will be presented; a slightly more general version will be stated and proved in Section 4. The object of interest is a sequence of random variables, taking values in Polish spaces. These range spaces may differ from variable to variable. For the i-th variable, let Ω_i be a Polish space, that is, a Borel subset of a compact metric space. Equip Ω_i with its Borel σ-field $\mathcal{F}_i$. Let $\Omega = \prod_{i=1}^{\infty} \Omega_i$ and $\mathcal{F} = \prod_{i=1}^{\infty} \mathcal{F}_i$. Let X_i be the i-th coordinate function on Ω. It is helpful to work with this concrete realization of the process. Often, X_i will be written for the i-th coordinate function on $\prod_{i=1}^{n} \Omega_i$.

The "sufficient statistic" T_n is a Borel mapping from $\prod_{i=1}^{n} \Omega_i$ to a Polish space W_n; write $\mathcal{B}_n$ for the Borel σ-field in W_n. In principle, T_n does not act on Ω; but $T_n(X_1, \ldots, X_n)$ does have domain Ω, because the coordinate functions X_i are defined on Ω. For each n and $t \in W_n$, let $Q_{n,t}$ be a probability on $\prod_{i=1}^{n} \mathcal{F}_i$ in $\prod_{i=1}^{n} \Omega_i$. It is assumed that $t \to Q_{n,t}$ is Borel. Define M_Q, the *partially exchangeable probabilities with respect to* Q_n *and* T_n, as the class of probabilities on $\mathscr{F}$ in Ω such that : for each n, given $T_n(X_1, \ldots, X_n) = t$, a regular conditional distribution for $X_1, \ldots, X_n$ is $Q_{n,t}$. Informally, $Q_{n,t}$ is the distribution of the data given that the sufficient statistic takes the value t. This does not depend on the parameters, i.e., it is the same for all $P \in M_Q$. More exactly, M_Q is the class of P for which this statement is true.

An example of this set-up is provided by coin-tossing. Then $\Omega_i = \{0, 1\}$, $\mathscr{F}_i$ is the discrete σ-field, $T_n(x_1, \ldots, x_n) = x_1 + \ldots + x_n$, $W_n = \{0, 1, \ldots, n\}$, $\mathscr{B}_n$ is the discrete σ-field in W_n. The relevant $Q_{n,t}$ assigns equal probability $1 \Big/ \binom{n}{t}$ to each of the $\binom{n}{t}$ sequences in $\prod_{i=1}^{n} \Omega_i$ with $T_n = t$. Then M_Q is precisely the class of exchangeable processes. De Finetti's theorem identifies the extreme points of M_Q, and states that any element of M_Q is a unique average of such extreme points. So de Finetti's theorem can be (with some effort) seen as a special case of Choquet's theorem. On this score, see Kendall (1963) or Phelps (1966). The object is to generalize these considerations.

It will be necessary to assume that the $Q_{n,t}$ fit together, as follows :

$$Q_{n,t}\{T_n = t\} = 1. \quad \ldots \quad (1.1)$$

If $T_n(x_1, \ldots, x_n) = T_n(x_1', \ldots, x_n')$, then
$T_{n+1}(x_1, \ldots, x_n, y) = T_{n+1}(x_n', \ldots, x_n', y)$ for all $y \in \Omega_{n+1}$. ... (1.2)

For each $s \in W_n$ and $t \in W_{n+1}$, relative to $Q_{n+1,t}$,
the kernel $Q_{n,s}$ is a regular conditional distribution
for $(X_1, \ldots, X_n)$ given $T_n(X_1, \ldots, X_n) = s$ and X_{n+1}. ... (1.3)

Related definitions have been given by Freedman (1962) and Bahadur (1954). For a discussion of such ideas, see Lauritzen (1974a). In the coin-tossing example, condition (1) is trivial; (2) is easy,

$$T_{n+1}(x_1, \ldots, x_n, y) = T_n(x_1, \ldots, x_n) + y.$$

Property (3) is almost as easy. For one thing, s and t determine $X_{n+1} = t - s = 0$ or 1. So it is vacuous to condition on X_{n+1}. Now put a string of $n+1$ symbols down in random order, where t are 1 and $n-t$ are 0. Given that among the first n exactly s are 1, the first n are still in random order. That is all (3) says.

Going back to the general case, it is easy to check that M_Q is convex. Theorem 1.1 below will characterize the extreme points of M_Q, and show that $P \in M_Q$ is a unique average of extreme points. The theorem is a bit abstract, so the characterization is indirect; in any concrete problem, some effort may be needed to identify the extreme points : see Sections 2 and 3. The extreme points are closely related to a certain σ-field $\hat{\Sigma}$, which will be called the *partially exchangeable* σ-field. Namely,

$$\hat{\Sigma} = \bigcap_{n=1}^{\infty} \hat{\Sigma}^{(n)}$$

where $\hat{\Sigma}^{(n)}$ is generated by $T_n(X_1, \ldots, X_n),\ X_{n+1},\ X_{n+2}, \ldots$.

In the coin-tossing example, $\hat{\Sigma}^{(n)}$ is the σ-field of measurable sets which are invariant under permutations of the first n coordinates. So $\hat{\Sigma}$ is the σ-field of measurable sets which are invariant under any finite permutation of coordinates. This is often called the *exchangeable* σ-field.

Theorem 1.1 : *Assume condition* (1.1–3). *There is a set* $E \in \hat{\Sigma}$ *with* $P(E) = 1$ *for all* $P \in M_Q$; *for each* $\omega \in E$, *the sequence of probabilities* $Q_{n, T_n[X_1(\omega), \ldots, X_n(\omega)]}$ *converges weak-star to a probability* $Q(\omega, \cdot) \in M_Q$. *This* $Q(\omega, \cdot)$ *is* $0-1$ *on* $\hat{\Sigma}$. *As* ω *ranges over* E, *the probabilities* $Q(\omega, \cdot)$ *range over the extreme points of* M_Q. *For any* $P \in M_Q$, *the kernel* $Q(\omega, A)$ *is a regular conditional* P*-distribution given* $\hat{\Sigma}$, *and*

$$P = \int_E Q(\omega, \cdot)\hat{P}(d\omega)$$

where $\hat{P}$ *is the restriction of* P *to* $\hat{\Sigma}$. *This representation is unique. In particular,* $P \in M_Q$ *is extreme iff* P *is* $0-1$ *on* $\hat{\Sigma}$; *equivalently, iff*

$$P\{\omega : Q(\omega, \cdot) = P\} = 1.$$

In the coin-tossing example, E is the set of ω for which

$$\frac{1}{n}\,[X_1(\omega)+\ldots+X_n(\omega)] \text{ converges as } n \to \infty.$$

Call the limit $\lambda(\omega)$. Then $Q(\omega, \cdot)$ makes the coordinates independent, each taking the value of 1 with probability $\lambda(\omega)$ and 0 with probability $1-\lambda(\omega)$. In short, $Q(\omega, \cdot)$ is coin-tossing, with a $\lambda(\omega)$-coin. The proof of the theorem is deferred to Section 4.

2. Examples and remarks

The first example is de Finetti's theorem for random variables with values in a Polish space $(S, \mathscr{A})$: an exchangeable sequence is a mixture of i.i.d. variables.

Example 2.1 : Let $\Omega_i = S$ and $\mathcal{F}_i = \mathcal{A}$ for all i. Let $T_n(x_1, ..., x_n) = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$, where δ_x is point mass at x : so T_n is the empirical distribution. Let $Q_{n,t}$ be uniform over the collection of all finite sequences $x_1, ..., x_n$ having the empirical distribution t. Typically, $Q_{n,t}$ will have n ! atoms, but it may have fewer, if there are repeats. Conditions (1.1–2–3) are easy to check, and it can be verified that M_Q consists of the exchangeable P, i.e., those invariant under permutations of finitely many coordinates. Let λ be a typical probability on $(S, \mathcal{A})$. Let P_λ on $(\Omega, \mathcal{F})$ make the coordinates independent, with common distribution λ. If $t_n \to \lambda$ weak-star, then $Q_{n, t_n} \to P_\lambda$ weak-star. If t_n fails to converge, so does Q_{n,t_n} : indeed, the Q_{n,t_n}-distribution of X_1 is just t_n. Let $\Lambda(\omega) = \lim_n T_n[X_1(\omega), ..., X_n(\omega)]$ on the set L where the limit exists. If $P \in M_Q$, then $P(L) = 1$, and for $\omega \in L$ the limit $Q(\omega, \cdot)$ is $P_{\Lambda(\omega)}$. Since $P_\lambda\{\Lambda = \lambda\} = 1$, it follows that $P_{\Lambda(\omega)}$ is the typical extreme point. So $E = L$ and

$$P = \int P_{\Lambda(\omega)} \hat{P}(d\omega).$$

This is de Finetti's theorem, in some disguise.

The next example characterizes mixtures of i.i.d. normal variables with mean 0 and variance σ^2 ($N(0, \sigma^2)$). The result dates back at least to Schoenberg (1938), who needed it to characterize metric spaces that can be isometrically imbedded into Banach spaces. It has been proved in Bayesian language by Freedman (1962b), and in the theory of Radon measures : Chapter 9 of Choquet (1969). For recent variants and a bibliography, see Eaton (1981) or Letac (1981). A generalization to covariance mixtures of normal vectors is given by Dawid (1977). For a generalization to l^p spaces, see Bretagnolle, et al. (1966) or Berman (1980). The result is often stated as follows : a sequence is a mixture of i.i.d. $N(0, \sigma^2)$ variables iff for each n the distribution of the first n variables is invariant under rotations. To put the result into the present framework, observe that the condition of rotational invariance is equivalent to the condition that given $(X_1^2+X_2^2+...+X_n^2)^{\frac{1}{2}} = t$, say, the conditional distribution of $(X_1, ..., X_n)$ should be uniform on the $(n-1)$-sphere of radius t in R^n. Denote this uniform distribution by $Q_{n,t}$.

Example 2.2 : Let Ω_i be the line, with the Borel σ-field $\mathcal{F}_i$. Let $T_n(x_1, ..., x_n) = (x_1^2+...+x_n^2)^{\frac{1}{2}}$. Let $0 \leqslant \sigma < \infty$, and let the probability P_σ on $(\Omega, \mathcal{F})$ make the coordinates independent, with common $N(0, \sigma^2)$ distribution. If $t_n/\sqrt{n} \to \sigma$, then $Q_{n,t_n} \to P_\sigma$ weak-star. This fact can be traced back to Maxwell, see page 134 of Everitt (1974). It has been rediscovered many times since. It follows that if $t_n/\sqrt{n}$ fails to converge, so does Q_{n,t_n}. For example, suppose $t_n/\sqrt{n} \rightarrow \infty$. The Q_{n,t_n}-law of X_1 coincides with the $Q_{n,\sqrt{n}}$-law of $X_1 \cdot t_n/\sqrt{n}$, and the mass drifts off to $\pm\infty$.

Let $\Lambda(\omega) = \lim_{n\rightarrow\infty} \frac{1}{n} [X_1(\omega)^2+\ldots+X_n(\omega)^2]^{\frac{1}{2}}$, on the set L where the limit exists. Again, $E \subset L$ so $P(L) = 1$ for $P \in M_Q$. For $\omega \in L$, the limit $Q(\omega, \cdot)$ is $P_{\Lambda(\omega)}$. Again, $P_\sigma\{\Lambda = \sigma\} = 1$, so $E = L$; the extreme $P \in M_Q$ are precisely the P_σ's, and for any $P \in M_Q$,

$$P = \int P_{\Lambda(\omega)}\hat{P}(d\omega).$$

Example 2.3 : When is $X_1, X_2, \ldots,$ a mixture of sequences of independent identically distributed normal variables, with common mean μ and variance σ^2 ? Both μ and σ^2 vary in the mixture. The necessary and sufficient condition is that given

$$U_n = X_1+\ldots+X_n$$

and

$$V_n = (X_1^2+\ldots+X_n^2)^{\frac{1}{2}},$$

the conditional distribution of $X_1, \ldots, X_n$ is uniform over the relevant $(n-2)$-sphere in R^n. The argument is as in the previous example. Alternatively, such mixtures may be characterized as the set of processes invariant under the group of orthogonal transformations of R^n that preserve the line $x_1 = x_2 = \ldots = x_n$; see Smith (1981) for further discussion.

Example 2.4 : When is $X_1, X_2, \ldots,$ a mixture of independent identically distributed normal variables with common mean μ and known variance σ^2 ? Only μ varies in the mixture. The necessary and sufficient condition is that X is exchangeable and that given $X_1+\ldots+X_n = t$, the joint distribution of $X_1, \ldots, X_n$ is normal, with $E(X_i) = t/n$, $\text{var}(X_i) = \sigma^2$, $\text{cov}(X_i, X_j) = -\sigma^2/n-1$. Thus the law of $X_1, \ldots, X_n$ is normal on the hyperplane $X_1+\ldots+X_n = t$. The argument is essentially the same as the argument for example 2.2.

Example 2.5 : When is $X_1, X_2, \ldots,$ a mixture over θ of sequences of independent uniform variables with range $[0, \theta]$? The necessary and sufficient condition is that given $M_n = \max(X_1 \ldots, X_n)$, the X_i's are independent and uniform over $[0, M_n]$, for $i = 1, \ldots, n$. The idea is that $M_n \uparrow \theta$, so the conditional law of $X_1, \ldots, X_k$ given M_n tends to the right limit.

Example 2.6 : Fix a sequence of positive constants c_i with $\Sigma c_i = \infty$. Let λ be a nonnegative parameter. Let $X_1, X_2, \ldots,$ be independent, X_i being

Poisson with parameter λc_i. Write P_λ for the law of $X_1, X_2, \ldots$. To characterize mixtures of P_λ's let

$$T_n = X_1 + \ldots + X_n.$$

These will be the sufficient statistics. Let W_n consist of the nonnegative integers, and for $t \in W_n$ let $Q_{n,t}$ be multinomial, with t trials and n cells, the i-th one having probability $c_i / \sum_{i=1}^{n} c_i$. Then the mixtures coincide with the class M_Q, and the P_λ are the extreme points. The argument is that Q_{n,t_n} converges iff $t_n/c_1+\ldots+c_n$ converges to a finite limit, say λ; and then the limit is P_λ. If $t_n/c_1+\ldots+c_n \to \infty$, the Q_{n,t_n}-law of X_1 is binomial, with success probability $c_1/c_1+\ldots+c_n$, and number of trials t_n, so the mass drifts off to ∞. Now one proceeds as usual, with $\Lambda(\omega) = \lim_{n\to\infty} T_n(w)/c_1+\ldots+c_n$ on the set L where the limit exists.

Example 2.7 : This is like example 2.6, but $\Sigma c_i < \infty$. The sufficient statistics are still T_n, with the same $Q_{n,t}$'s. The twist is that the P_λ's above are no longer extreme in M_Q. Instead, the extreme points are the multinomials, with t balls being dropped into a countably infinite number of boxes; the balls are independent, and each ball drops into box i with probability $c_i/\Sigma c_i$; let Q_t be the distribution of the box counts; then Q_t is a typical extreme point. The idea is that Q_{n,t_n} converges iff t_n converges to a finite limit, and this will be t.

Remark 2.1 : Mixtures of Markov chains can be characterized by using the transition counts as sufficient statistics as in Freedman (1962a), Diaconis and Freedman (1980a). These results can also be derived as a special case of Theorem 1.1.

Remark 2.2 : We have not treated a host of more or less ob ious consequences of the representation. These include genera'izations of the Hewitt-Savage zero-one law and theorems involving limiting theory of partially exchangeable variables. The bibliography lists a number of papers which give detailed discussion of special cases.

Remark 2.3 : In principle, the same kind of reasoning applies to characterize mixtures of P_λ, where relative to P_λ, the X_i are independent generalized exponentials

$$P_\lambda\{X_i = j\} = c(i, \lambda)e^{-f(j)g(\lambda)}\, \mu_i\{j\}.$$

Applications include the binomial and negative binomial. See Freedman (1962a) for details.

In recent work, Berg and Ressel (1978) and Ressel (1983) treat examples where the sufficient statistics are "sums" with values in a semigroup. For many cases, they can identify the extreme points with characters in the dual semigroup.

The examples suggest a simple scheme for generating others. Take any standard parametric family P_λ; let $X_1, X_2, \dots$ be i.i.d. P_λ. Consider the usual sufficient statistic for λ, $T_n(X_1, \dots, X_n)$. Let $Q_{n,t}$ be the conditional distribution of $X_1, X_2, \dots, X_n$ given $T_n = t$. The T_n and $Q_{n,t}$ define a convex set M_Q as in Theorem 1. Sometimes the extreme points of M_Q are precisely the family P_λ. Examples (3.5–3.7) show that this scheme can fail. A positive result for integer valued exponential families with $T_n = X_1 + \dots + X_n$ is given in Section 3. We do not know a version for continuous exponential families.

Remark 2.4 : *Invariance and partial exchangeability.* Consider again de Finetti's theorem for exchangeable random variables. The generalizations discussed here depend on the formulation in terms of the conditional distribution given a statistic. Another way to generalize is to consider processes invariant under groups other than the permutation group. There is a widely known representation theory for invariant measures. Roughly, every invariant probability is a unique mixture of extreme invariant probabilities : the extreme points are characterized as ergodic, zero-one on the σ-field of invariant sets. Details may be found in Farrell (1962), Varadarajan (1963), Phelps (1966), or Maitra (1977).

There does not seem to be any theory that says when the extreme points may be regarded as a family of measures smoothly parametrized by a low dimensional manifold. Consider examples 2.1–2.3 from the invariant point of view :

In Example 2.1 (de Finetti's theorem for Polish space valued variables) the set M_Q is the set of all probabilities invariant under the permutation group. Here if Ω_i is infinite, the extreme points are infinite dimensional; while if the basic space Ω_i is finite, the extreme points are finite dimensional. Thus "finite dimensional extreme points" are not simply a property of the group involved. Examples 2.2 and 2.3 (mixtures of normals) can be characterized by invariance under the orthogonal group and a subgroup of the orthogonal group. We do not know a group theoretic characterization for location mixtures of normals with known scale.

A further example is provided by a theorem of David Aldous (1981b). Aldous considered the problem of an array X_{ij} with joint distributions invariant under permuting rows and columns. He determined the extreme points and a representation theorem. Hoover (1981a) contains further specification of the extreme points. Diaconis and Freedman (1981) contains an application to a psychology problem. Here, even if the X_{ij} only take values zero and one, the extreme points are infinite dimensional. This problem can be thought of as a sequence of random variables invariant under a subgroup of the infinite permutation group.

Remark 2.5 : *Finite versions of the theorem.* Theorem 1.1 characterizes the extreme points of infinite partially exchangeable sequences. For exchangeable sequences, there have been a number of results that imply that a finite exchangeable sequence is almost a mixture of coin-tossing. For details and references to the work of de Finetti, Kendall, and others on this problem, see Diaconis and Freedman (1980c). In Diaconis and Freedman (1983) we prove that if $X_1, ..., X_n$ is orthogonally invariant then, for $k < n$, the law of $X_1, ..., X_k$ is almost a scale mixture of i.i.d. $N(0, \sigma^2)$ variables in variation distance with error smaller than $2k/n$. Similar results hold for mixtures of geometric, Poisson, and exponential variables. A general theory is lacking. Recently there has been interesting work showing that finite exchangeable sequences can be characterized as mixtures of coin-tossing with a possibly negative mixing measure being allowed. A result of P. A. Meyer is given in Delacherie-Meyer (1975, p. 48–53). See also Jaynes (1982).

Zaman (1981) considers zero-one valued processes and $T_n(x_1, x_2, ..., x_n) = (t_{ij})$ where t_{ij} is the number of i to j transitions in $(x_1, ..., x_n)$, $i, j = 0, 1$. The $Q_{n,t}$ measures are taken as uniform over all binary sequences of length n with $T_n = t$. Zaman shows that if $X_1, X_2, ..., X_n$ is a stationary partially exchangeable sequence of length n which can be extended to a partially exchangeable sequence of length $n+k$ then, in variation distance, $X_1, X_2, ..., X_n$ is almost a mixture of stationary Markov chains, the error tending to zero like $(\log k)/k$.

3. Integer-valued processes with sums as sufficient statistics

This section characterizes integer-valued processes $X_1, X_2, ...,$ with the partial sums $T_n = X_1 + ... + X_n$ as sufficient statistics. The extreme points turn out to be sequences of independent random variables with common generalized exponential distribution. We will write $q_{n,k}$ for a proposed conditional distribution of $(X_1, ..., X_n)$ given $X_1 + ... + X_n = k$; let M_q be the set of processes $X_1, X_2, ...,$ such that $q_{n,k}$ is in fact the conditional distribution of $X_1, ..., X_n$ given $X_1 + ... + X_n = k$, Some three examples will give the flavor of the theory to be developed here.

Example 3.1 : Let $q_{n,k}$ be the uniform distribution over all nonnegative n-tuples of integers whose sum is k. Then the extreme points of M_q are the sequences of independent random variables with common geometric distribution.

Example 3.2 : Let $q_{n,k}$ be the multinomial distribution on n-tuples of nonnegative integer whose sum is k, with uniform success probabilities $(1/n, \ldots, 1/n)$. Then the extreme points of M_q are the sequence of independent random variables with common Poisson distribution.

The geometric distribution and the Poisson are exponential. The next example patches the two previous examples together.

Example 3.3 : If $k \geqslant 0$, let $q_{n,k}$ be the multinomial distribution on n-tuples of integers whose sum is k. If $k \leqslant 0$, let $q_{n,k}$ be the uniform distribution over all n-tuples of nonpositive integers whose sum is k. If $k = 0$, by either definition, $q_{n,k}$ sits on the single n-tuple $(0, \ldots, 0)$. Then there are two kinds of extreme points of M_q :

- Sequences of independent nonnegative random variables all with common Poisson distribution.
- Sequences of independent non-positive random variables which have, when their sign is reversed, a common geometric distribution.

In the general case, the situation may be described as follows : q determines a partition of the integers into intervals I_j, disjoint except for endpoints, and reference measures μ_j supported by I_j. An extreme point of M_q is a sequence of independent random variables, with common distribution an exponential, through some μ_j.

Implicit in the proof is a version of the Koopman-Pitman-Darmois theorem for integer valued random variables. Our results have some overlap with results of Lauritzen (1975), (1982; Chapter 3) who develops, in a different language, a theory of "generalized exponential families" on a discrete set using the language of semi-groups.

Let I be a set of integers of the form $I = \{n \in Z : a \leqslant n \leqslant b\}$ where a and b are allowed to be infinite. Let μ be a nonnegative measure on I. A probability p on I is *expo·ential through* μ if there is a real number λ and a constant c such that

$$p(n) = ce^{\lambda n}\mu(n) \text{ for } n \in I.$$

Of course, if I is all the integers, and μ grows rapidly at $\pm\infty$, there may be no exponentials at all through μ. Also by definition, an "exponential μ-sequence" is a sequence of independent random variables, with a common distribution which is exponential through μ.

For each n, let D_n be a subset of the integers, and for $t \in D_n$ let $q_{n,t}$ be a probability on n-tuples of integers. Suppose $q_{n,t}$ is exchangeable, so all permutations of any given n-tuple will have the same $q_{n,t}$-probability. It will always be assumed that the D's and q's are consistent, in the following sense : if $t_{n+1} \in D_{n+1}$ and $q_{n+1,t_{n+1}}(x_1, \ldots, x_n, x_{n+1}) > 0$, then

$$x_1+\ldots+x_n+x_{n+1} = t_{n+1}$$
$$x_1+\ldots+x_n = t_n \in D_n.$$

Furthermore,

$$q_{n+1,t_{n+1}}(x_1, \ldots, x_n \mid x_1+\ldots+x_n = t_n) = q_{n,t_n}(x_1, \ldots, x_n).$$

Trivially, t_{n+1} and t_n determine $x_{n+1} = t_{n+1}-t_n$, so there is no need to condition on x_{n+1} : compare (1.3). The D_n and $q_{n,t}$ will be considered as given.

A sequence of integer-valued random variables $X_1, X_2, \ldots$, will be called "q-able" provided that for all n,

$$P(S_n = t) > 0 \text{ implies } t \in D_n$$

and

$$P(X_1 = x_1, \ldots, X_n = x_n \mid S_n = t) = q_{n,t}(x_1, \ldots, x_n)$$

where $S_n = X_1+\ldots+X_n$. Informally, the partial sums are sufficient statistics; at stage n, given $S_n = t$, the conditional distribution for the data is $q_{n,t}$. Write M_q for the class of q-able measures.

It will be clear that an exponential-μ sequence is q-able, with q depending on μ. Taking a mixture of such sequences over various parameters λ, preserves the q-able structure, and even q itself provided μ is held constant. It is even possible to mix over μ's, to a certain extent : the intervals of support for the various μ's must be disjoint except for endpoints. The next theorems establish the converse : a q-able sequence must be a mixture of exponential-μ sequences, where it is feasible to mix over μ itself to the extent indicated above.

Suppose now D and q are given and consistent. Consider a q-able process. Theorem 1.1 can be used, for the q-able process has law $P \in M_q$. Consider the extreme points $Q(\omega, \cdot)$ for $\omega \in E$. On the one hand, $Q(\omega, \cdot) \in M_q$ by the theorem. On the other, $Q(\omega, \cdot)$ makes the coordinate variables independent and identically distributed. This is de Finetti's theorem, for $Q(\omega, A)$ is a regular conditional distribution given $\hat{\Sigma}$. A q-able process with law P is then a mixture $\int Q(\omega, \cdot)\hat{P}(d\omega)$ of sequences of i.i.d. variables which are q-able. Such sequences are exponential μ-sequences for certain μ, as the next theorem shows.

Theorem 3.1 : *Given D and q consistent, there is determined a finite or countable collection $\mathcal{C}$ of intervals $\{I, J, \ldots\}$, pairwise disjoint except for end-points, and for each $I \in \mathcal{C}$ a finite or infinite measure μ_I concentrating on I, with the following property. If $X_1, X_2, \ldots$, are i.i.d. and q-able, then for some $I \in \mathcal{C}$, the common law of the X_n is exponential through μ_I.*

The following points may be noted :

(i) μ_I need not assign positive mass to all the elements of I.

(ii) If I is infinite in both directions, then $\mathcal{C} = \{I\}$; there need not be any exponential law through μ_I, in which case there are no q-able X's.

The theorem is an immediate consequence of the following proposition. As discussed at the end of this section, the proposition is a version of the Pitman-Koopman-Darmois theorem for discrete variables.

Proposition 3.1 : *Suppose the random variables X_n are i.i.d. and q-able and likewise for X'_n.*

(a) *There are only three possibilities :*
- *either ess.sup. $X_n \leqslant$ ess.inf. X'_n*
- *or ess.sup. $X'_n \leqslant$ ess.inf. X_n*
- *or the laws of X_n and X'_n have the same support,*

(b) *In the last case, the law of X'_n is exponential through the law of X_n.*

The proof of Proposition 3.1 depends on the following lemmas.

Lemma 3.1 : *Let $i < j < k$ be integers. Then there are positive integers a, b, c such that*

$$ai+ck = bj$$
$$a+c = b.$$

Proof : Choose a and c so that

$$c(k-j) = a(j-i),$$

For example, $c = (j-i)$ and $a = (k-j)$. Then set $b = a+c$. □

Lemma 3.2 : *Let p' and p be two probabilities on the integers. with the same support S. Suppose that for all distinct $i, j, k \in S$*

$$\frac{\lambda_k-\lambda_i}{k-i} = \frac{\lambda_j-\lambda_i}{j-i}, \qquad \ldots (3.1)$$

where $\lambda_i = \log p'_i/p_i.$

Then p' is exponential through p. Conversely, if p' is exponential through p, then (3.1) *holds.*

Proof: The ratios $(\lambda_j-\lambda_i)/(j-i)$ all have a common value, call it λ. Then fixing $i_0 \in S$,

$$\lambda_j = \lambda_{i_0} - \lambda i_0 + \lambda j$$

so

$$p_j = Ce^{\lambda j} p'_j$$

with

$$C = \exp[\lambda_{i_0} - \lambda i_0].$$

The converse is clear. □

Lemma 3.3: *Let $i < j < k$ be integers, let $\lambda_i, \lambda_j, \lambda_k$ be real numbers. Suppose that for integer α, β, γ, conditions* (3.2–3) *imply* (3.4):

$$\alpha+\beta+\gamma = 0 \quad \ldots \quad (3.2)$$

$$\alpha i+\beta j+\gamma k = 0 \quad \ldots \quad (3.3)$$

$$\alpha\lambda_i+\beta\lambda_j+\gamma\lambda_k = 0. \quad \ldots \quad (3.4)$$

Then equation (3.1) *holds.*

Note: In the application, the λ's will be as in Lemma 3.2, but this is irrelevant here.

Proof: As in Lemma 3.1, select β and γ to be relatively prime integers satisfying

$$\beta(j-i)+\gamma(k-i) = 0. \quad \ldots \quad (3.5)$$

Set $\alpha = -(\beta+\gamma)$. Clearly, (3.2) and (3.3) hold, so (3.4) holds. Substitute $\alpha = -(\beta+\gamma)$ in (3.4):

$$\beta(\lambda_j-\lambda_i)+\gamma(\lambda_k-\lambda_i) = 0. \quad \ldots \quad (3.6)$$

But (3.5) entails

$$\beta = -\gamma(k-i)/(j-i)$$

and in particular, $\gamma \neq 0$. Substitute into (3.6) and cancel γ to get (3.1). □

Lemma 3.4: *Let $i < j < k$ be integers; let $\lambda_i, \lambda_j, \lambda_k$ be real numbers. Let a, b, c, a', b', c' be nonnegative integers. Suppose conditions* (3.7–8) *imply* (3.9):

$$a+b+c = a'+b'+c' \quad \ldots \quad (3.7)$$

$$ai+bj+ck = a'i+b'j+c'k \quad \ldots \quad (3.8)$$

$$a\lambda_i+b\lambda_j+c\lambda_k = a'\lambda_i+b'\lambda_j+c'\lambda_k. \quad \ldots \quad (3.9)$$

Then equation (3.1) *must hold.*

Proof: The idea is to use Lemma 3.3. with $\alpha = a - a'$, etc. More specifically, let α, β, γ be a triple of signed integers satisfying (3.2) and (3.3). Construct a and a' as follows :

$$\text{if } \alpha \geqslant 0 \text{ then } a = \alpha \text{ and } a' = 0$$
$$\text{if } \alpha < 0 \text{ then } a = 0 \text{ and } a' = -\alpha.$$

Likewise for b, b' and c, c'. Then (3.7) and (3.8) hold. So (3.9) holds, and this is (3.4). So (3.1) holds. □

Proof of Proposition 3.1 : Let

$$S_n = X_1 + \ldots + X_n, \qquad S'_n = X'_1 + \ldots + X'_n$$
$$p_i = P(X_n = i), \qquad p'_n = P(X' = i).$$

Part (a). Let $i < j < k$. Suppose p_i, p'_j, and p_k are all positive. It is claimed that p'_i, p_j, and p'_k are positive too. See Table 1.

TABLE 1. ENTRIES MARKED "?" ARE TO BE PROVED POSITIVE

	p	p'
i	+	?
j	?	+
k	+	?

To prove the claim, apply Lemma 3.1 to i, j, k. Let $t = bj$ and $n = b$; now $n - a = b - a = c$. Since $t = ai + ck$,

$$P(S_n = t) \geqslant P(X_1 = \ldots = X_a = i \text{ and } X_{a+1} = \ldots = X_n = k) = p_i^a p_k^c > 0,$$

clearly, $$P(S'_n = t) \geqslant P(X_1 = \ldots = X_n = j) = p_j^b > 0.$$

Especially, $t \in D_n$. Since X and X' are q-able,

$$p_j^b / P(S_n = t) = q_{nt}(j, \ldots, j)$$
$$= p_j'^b / P(S'_n = t) > 0$$

so $p_j > 0$ as claimed. Likewise,

$$p_i'^a p_k'^c / P(S'_n = t) = p_i^a p_k^c / P(S_n = t) > 0,$$

so $p'_i > 0$ and $p'_k > 0$, as claimed. Part (a) of the proposition now follows, as a moment's thought will show.

Part (b). The idea is to apply Lemma 3.2; equation (3.1) holds by Lemma 3.4. More specifically, let $i < j < k$ be in the common support of the law of X_n and X'_n. Let a, b, c, a', b', c' be nonnegative integers such that (3.7) and (3.8) hold :

$$a+b+c = a'+b'+c' = n \text{ say}$$
$$ai+bj+ck = a'i+b'j+c'k = t \text{ say.}$$

Then
$$P(S_n = t) \geqslant p_i^a\, p_j^b\, p_k^c > 0$$
$$P(S'_n = t) \geqslant p_i'^a\, p_j'^b\, p_k'^c > 0$$

so $t \in D_n$. Because X and X' are both q-able.

$$p_i^a\, p_j^b\, p_k^c / P(S_n = t) = p_i'^a\, p_j'^b\, p_k'^c / P(S'_n = t).$$

Define λ_i as in Lemma 3.2 and take logs :

$$a\lambda_i + b\lambda_j + c\lambda_k = \log P(S'_n = t) - \log P(S_n = t).$$

Likewise for a', b', c', proving (3.9). Then (3.1) holds, and Lemma 3.2 proves claim (b).

Remark : These results seem to extend to the case where the sufficient statistics are $\sum_{i=1}^{n} h(X_i)$, with h an integer-valued function. The relevant "exponential" distributions are of the form

$$c(\lambda)\, e^{\lambda h(x)}\, \mu\{x\}.$$

They may also extend to the case of vectors of integers, but the relationship among the cases (Proposition 3.1a) is not clear; also, Proposition 3.1b needs attention in the vector case. Both Martin-Löf (1974) and Lauritzen (1982) have some results for the vector case.

Given D_n and q_{nt} consistent, when does there exist a q-able i.i.d. sequence ? We do not have a neat answer. and the question is nontrivial, as the following examples show.

Example 3.4 : There are D and q consistent, but no q-able processes : Let $X_{N1}, \ldots, X_{NN}$ have values $N-1$ and one value $-(N-1)^2$, placed in random order. Let $S_{Nn} = X_{N1} + \ldots + X_{Nn}$ for $n = 1, \ldots, N$. Let

$$D_{Nn} = \{t : P(S_{Nn} = t) > 0\}$$

and for $t \in D_{Nn}$, let $q_{N,n,t}$ be the conditional distribution of $X_{N1}, \ldots, X_{Nn}$ given $S_{Nn} = t$. Fix n. The D_{Nn} are pairwise disjoint, for $N = n, n+1, \ldots$. Indeed, $D_{nn} = \{0\}$ and for N larger, D_{Nn} contains the two values $\{(N-1)n, -(N-1)(N-n)\}$. Let $D_n = \bigcup_{N=n}^{\infty} D_{Nn}$; if $t \in D_n$, then $t \in D_{Nn}$ for a unique N; let $q_{nt} = q_{Nnt}$. These D's and q's are consistent. Indeed, fix $t \in D_{N,n+1}$ with $N \geqslant n+1$. Suppose $q_{N,n+1,t}$ assigns positive mass to $x_1, \ldots, x_n, x_{n+1}$. Then there must be positive probability that

$$X_{N1} = x_1, \ldots, X_{Nn} = x_n, X_{N,n+1} = x_{n+1}.$$

There are only three possible cases :

Case 1 : $x_1 = \ldots = x_n = x_{n+1} = N-1$

Case 2 : $x_1 = \ldots = x_n = N-1$ and $x_{n+1} = -(N-1)^2$

Case 3 : $x_1 = \ldots = x_n = N-1$, except for one i with $x_i = -(N-1)^2$; and $x_{n+1} = N-1$.

In all three cases, $s = x_1 + \ldots + x_n \in D_{Nn}$; and the $q_{N,n+1,t}$ -distribution of $X_1, \ldots, X_n$ given $X_1 + \ldots + X_n = s$ is $q_{N,n,s}$.

However, there is no q-able process. For if there were, for some sequence t_n with $t_n \in D_n$, the q_{n,t_n}-law of X_1 would have to converge, by Theorem 1.1. However, all the mass escapes to $\pm\infty$. Indeed, suppose $t_n \in D_{Nn}$ with $N \geqslant n$. Then the only q_{n,t_n}-possible values for X_1 are $N-1$ and $-(N-1)^2$.

Theorem 3.1 and Proposition 3.1 are discrete versions of the Koopman-Pitman-Darmois theorem. The usual version of this theorem says that if a suitably smooth family $P_\lambda(dx)$ admits the sum as a sufficient statistic, then the family is an exponential family. A clean version of this theorem is in Hipp (1974), which also has useful references to previous work.

Proposition 3.1 implies that if a family of probabilities P_λ on the integers have common support, and if for each n the sum is a sufficient statistic, then the family is exponential. Anderson (1970) has this result and results for some sufficient statistics other than sums. The followng example shows that if h is a real valued function and $\sum_{i=1}^{n} h(X_i)$ is sufficient the family may not be exponential.

Example 3.5 : Let $\mathscr{N}$ be the family of all probabilities whose support is the integers. Let $T_n(X_1, \ldots, X_n) = \sum_{i=1}^{n} h(X_i)$, where h is the exponential function. Then T_n is sufficient for $\mathscr{N}$. Indeed, e is transcendental. Therefore, knowing T_n is the same as knowing the order statistics, and these are sufficient.

The usual versions of the Koopman-Pitman-Darmois Theorem imply that if the sum is sufficient for samples of size $n = 2$, then the family is exponential. The discrete version given here requires a condition on samples of every size. The following example shows that no fixed, finite value of n will do in the discrete case.

Example 3.6 : Let N be a positive integer. Let $\mathscr{N}_N$ be the family of all probabilities supported on the powers of N, namely $\{1, N, N^2, N^3, \ldots\}$. For $n < N$, the sum of the observations from a sample of size n are sufficient for $\mathscr{N}_N$ since the order statistics are recoverable from the sum.

The final example shows what happens if integer valued statistics more general than sums are admitted. The example is a naturally occurring complete exponential family with the product of the observations as sufficient statistics. The family is one-dimensional, but the associated extreme points are infinite dimensional.

Example 3.7 : *The zeta density.* For $1 < \lambda < \infty$, let

$$p_\lambda(n) = \frac{1}{\zeta(\lambda)n^\lambda}, \qquad n = 1, 2, \ldots,$$

where $\zeta(\lambda) = \Sigma\, 1/n^\lambda$ is Riemann's zeta function. The sufficient statistic for λ based on a sample of size n is the product T_n of the observations. The conditional distribution $q_{n,k}$ is uniform over all positive integer n-tuples with product equal to k.

Let M_Q be the family of all processes $X_1, X_2, \ldots$ such that $q_{n,k}$ is the conditional distribution of $X_1, \ldots, X_n$ given T_n. We now describe the general extreme point of M_Q. For prime p, let $\phi(p)$ be an arbitrary function with $0 < \phi(p) < 1$ and $\Sigma_p \phi(p) < \infty$. Extend ϕ to a multiplicative function on the positive integers as follows

$$\phi(1) = 1, \quad \phi(n) = \Pi_{p \mid n}\, \phi(p).$$

The product is over all prime divisors of n, so $\phi(12) = \phi(2)\phi(2)\phi(3)$. Expanding formally, using unique factorization,

$$\Sigma_n \phi(n) = \Pi_p (1-\phi(p));$$

the product converges because the sum of $\phi(p)$ is finite. Thus ϕ can be normalized to a probability $\hat{\phi}$ on $\{1, 2, 3, \ldots\}$.

It will now be shown that the extreme points of M_Q are the sequence of i.i.d. random variables with common distribution $\hat{\phi}$. To begin with, if $X_1, X_2, \ldots$ are i.i.d. according to $\hat{\phi}$, then the process $\{X_i\} \in M_Q$ because ϕ is multiplicative. The i.i.d. process is extreme in M_Q because it is extreme in the larger class of all exchangeable probabilities.

To show that all the extreme points of M_q arise in this way, observe first that de Finetti's theorem implies that any extreme point consists of a sequence of independent random variables with common distributions ψ; the argument is completed by showing that $\psi = \hat{\phi}$ for some ϕ. Consider X_1 and X_2 independent with common distribution ψ. Consider pairs of integers (i, j) and (i', j') with $ij = i'j' = 1 \cdot k$ say. Because the product is sufficient, $\psi(i)\psi(j) = \psi(i')\psi(j') = \psi(1)\psi(k)$. So $\psi(1) \neq 0$. Define $\phi(i) = \psi(1)^{-1}\psi(i)$ and observe $\phi(i \cdot j) = \phi(i)\phi(j)$ so $\phi(n) = \Pi_{p \mid n} \phi(p)$. It follows that ϕ is between 0 and 1 and that $\phi(p)$ sums as required.

Remarks : The zeta distribution is useful in probabilistic number theory because, for m and n relatively prime, $P_\lambda(m \text{ and } n \text{ divide } X) = P_\lambda(m \text{ divides } X)$ $p_\lambda(n \text{ divides } X)$. Diaconis (1980) gives applications.

In the continuous version of this example one considers the family with density $p_\lambda(dx) = \dfrac{1}{(\lambda-1)X^\lambda}\, dx$ on $[1, \infty)$ with $1 < \lambda < \infty$. Taking the products as sufficient statistics T_n and the conditional distribution given $T_n = t$ as $Q_{n,t}$, the family M_Q can be shown to have only i.i.d. sequences with law $p_\lambda(dx)$ as extreme points.

4. Mathematical appendix

Consider a sequence of Polish spaces $\mathscr{X}_n$ with their Borel σ-fields $\mathscr{A}_n$. Let p_n be a Borel in mapping of $\mathscr{X}_n$ onto $\mathscr{X}_{n-1}$, for $n \geqslant 2$. In Theorem 1.1,

$$\mathscr{X}_n = \prod_{i=1}^{n} \Omega_i \quad \text{and} \quad \mathscr{A}_n = \prod_{i=1}^{n} \mathscr{Q}_i$$

$$p_n(x_1, \ldots, x_{n-1}, x_n) = (x_1, \ldots, x_{n-1}).$$

Let W_n be a Polish space with Borel σ-field $\mathscr{B}_n$; let T_n be a Borel mapping from $\mathscr{X}_n$ into W_n. For each n and t, suppose $Q_{n,t}$ is a probability on $(\mathscr{X}_n, \mathscr{A}_n)$. with $t \to Q_{n,t}$ Borel. Suppose

$$Q_{n,t}\{T_n = t\} = 1. \qquad \ldots \quad (4.1)$$

For each n and $t \in W_{n+1}$, relative to $Q_{n+1,t}$, a regular conditional distribution for p_{n+1} given $T_n \circ p_{n+1} = s$ is $Q_{n,s}$. ... (4.2)

This condition is slightly weaker than (1.3), and the definition of the partially exchangeable σ-field will have to be revised slightly, in consequence.

Consider now the projective limit $\mathscr{X}$ of $\mathscr{X}_n$, namely, all infinite sequences $(x_1, x_2, \ldots)$ with $x_n \in \mathscr{X}_n$ and $p_n(x_n) = x_{n-1}$. This is a Borel subset of the product space $\prod_{i=1}^{\infty} \mathscr{X}_i$. Equip $\mathscr{X}$ with the Borel σ-field $\mathscr{A}$. Let ξ_n be the n-th coordinate function on $\mathscr{X}$, so $p_n(\xi_n) = \xi_{n-1}$. In Theorem 1.1, $\mathscr{X} = \Omega$ and $\mathscr{A} = \mathscr{F}$, but

$$\xi_n = (X_1, \ldots, X_n).$$

Introduce M_Q, the class of partially exchangeable probabilities, as follows: $P \in M_Q$ iff for all n, relative to P, the kernel $Q_{n,t}$ is a regular conditional distribution for ξ_n given $T_n(\xi_n) = t$.

The partially exchangeable σ-field is now defined as follows:

$$\Sigma = \bigcap_{n=1}^{\infty} \Sigma^{(n)}$$

where $\Sigma^{(n)}$ is generated by $T_n(\xi_n)$, $T_{n+1}(\xi_{n+1})$, ...

In short, Σ is the tail σ-field of the $T_n(\xi_n)$'s. Set-theoretically, Σ is usually smaller than the $\hat{\Sigma}$ of Theorem 1.1. In the presence of condition (1.3), however, the two σ-fields are measure-theoretically the same. This will be proved as Lemma 4 below.

Theorem 4.1: *Assume conditions* (4.1–2). *There is a set* $\mathscr{X}_E \in \Sigma$ *with* $P(\mathscr{X}_E) = 1$ *for all* $P \in M_Q$; *for each* $x \in \mathscr{X}_E$, *the sequence of probabilities* $Q_{n,T_n[\xi_n(x)]}$ *converges weak-star to a probability* $Q(x, \cdot) \in M_Q$. *This* $Q(x, \cdot)$ *is* $0-1$ *on* Σ. *As* x *ranges over* E, *the probabilities* $Q(x, \cdot)$ *range over the extreme points of* M_Q. *For any* $P \in M_Q$, *the kernel* $Q(x, A)$ *is a regular conditional distribution for* P *given* Σ, *and*

$$P = \int_{\mathscr{X}_E} Q(x, \cdot) d\hat{P}$$

where $\hat{P}$ *is the restriction of* P *to* Σ. *This representation is unique. In particular,* $P \in M_Q$ *is extreme iff* P *is* $0-1$ *on* Σ; *equivalently, iff*

$$P\{x : Q(x, \cdot) = P\} = 1.$$

The proof of the theorem is presented as a series of lemmas. First, some general facts about conditioning will be developed; routine proofs are omitted. Let X and Y be measurable mappings from $(\Omega, \mathscr{F}, P)$ to $(\Omega_X, \mathscr{F}_X)$ and $(\Omega_Y, \mathscr{F}_Y)$ respectively; so PY^{-1} is the distribution of Y, and is a probability on $(\Omega_Y, \mathscr{F}_Y)$. As usual, $Q_y(A)$ is a *regular conditional distribution* (r.c.d.) for X given $Y = y$ iff:

· for each $y \in \Omega_Y$, $Q_y(\cdot)$ is a probability on $(\Omega_X, \mathscr{F}_X)$;

· for each $A \in \mathscr{F}_X$, $y \to Q_y(A)$ is $\mathscr{F}_Y$-measurable;

· $Q_Y(A)$ is a version of $P(X \in A \mid Y)$.

In the first lemma, let f be a measurable mapping from $(\Omega_X, \mathcal{F}_X)$ to $(\Omega_f, \mathcal{F}_f)$.

Lemma 4.1 : *If Q_y is an r.c.d. for X given $Y = y$, then $Q_y f^{-1}$ is an r.c.d. for $f(X)$ given $Y = y$.*

Lemma 4.2 : *Suppose Q_y is an r.c.d. for X given $Y = y$. Let F be a nonnegative $\mathcal{F}_X \times \mathcal{F}_Y$-measurable function on $\Omega_X \times \Omega_Y$. Then*

$$E\{F(X, Y)\} = \int_{\Omega_Y} \int_{\Omega_X} F(x, y) Q_y(dx) P Y^{-1}(dy).$$

The next lemma may be intuitively obvious, but the proof is somewhat technical. The mapping Y and its range Ω_Y do not appear in the statement; y is used as the typical value of a function g.

Lemma 4.3 : *Let f, g, and h be measurable mappings from $(\Omega_X, \mathcal{F}_X)$ to $(\Omega_f, \mathcal{F}_f)$, $(\Omega_g, \mathcal{F}_g)$, and $(\Omega_h, \mathcal{F}_h)$ respectively. Let Q be an r.c.d. for X given $g(X) = y$. Thus, $y \in \Omega_g$ and Q_y is a probability on $(\Omega_X, \mathcal{F}_X)$. For each $y \in \Omega_g$, let Q_{yz} be an r.c.d. for f given $h = z$, relative to Q_y. Suppose $(y, z) \to Q_{yz}(A)$ is measurable for each $A \in \mathcal{F}_f$. Then Q_{yz} is an r.c.d. for $f(X)$ given $g(X) = y$ and $h(X) = z$.*

In other words : conditioning first that $g(X) = y$ and second that $h(X) = z$ is the same as conditioning simultaneously that $g(X) = y$ and $h(X) = z$.

Proof : Let $A \in \mathcal{F}_f$, $B \in \mathcal{F}_g$, and $C \in \mathcal{F}_h$. What must be shown is that

$$P\{f(X) \in A \text{ and } g(X) \in B \text{ and } h(X) \in C\} = \int 1_B[g(X)] 1_C[h(X)] Q_{g(X)h(X)}(A) dP. \quad \ldots \quad (4.3)$$

The left side of (4.3) is

$$P\{X \in (f^{-1}A \cap h^{-1}C) \text{ and } g(X) \in B\} = \int Q_y(f^{-1}A \cap h^{-1}C) 1_B(y) P g(X)^{-1}(dy).$$

Now bring in Q_{yz}; for $y \in \Omega_g$ and $z \in \Omega_h$ this is a probability on Ω_f, and

$$Q_y(f^{-1}A \cap h^{-1}C) = \int_{\Omega_h} Q_{yz}(A) 1_C(z) Q_y h^{-1}(dz).$$

So the left side of (4.3) equals

$$\int_{\Omega_g} \int_{\Omega_h} Q_{yz}(A) 1_B(y) 1_C(z) Q_y h^{-1}(dz) P g(X)^{-1}(dy).$$

The "change of variables" formula can be used to express the inner integral as an integral with respect to $Q_y(dx)$ over $x \in \Omega_X$. So the left side of (4.3) equals

$$\int_{\Omega_g} \int_{\Omega_X} Q_{yh(x)}(A) 1_B(y) 1_C[h(x)] Q_y(dx) P g(X)^{-1}(dy).$$

To evaluate this last expression, use lemma 4.2 with $g(X)$ for Y and

$$F(x, y) = Q_{yh(x)}(A)1_B(y)1_C[h(x)].$$

The left side of (4.3) is then equal to the right side. □

Return now to the setting of Theorem 4.1. It will be useful to define for $m > n$ a Borel mapping $p_{m,n}$ of $\mathscr{X}_m$ onto $\mathscr{X}_n$, as follows :

$$p_{m,n} = p_{n+1} \circ p_{n+2} \circ \dots \circ p_m.$$

By convention, $p_{m+1,m} = p_m$ and $p_{m,m}$ is the identity. The next lemma is somewhat technical, but it is one of the key steps in the proof.

Lemma 4.4 : *Let* $m > n$, *and* $t \in W_n$. *Relative to* $Q_{m,t}$, *an r.c.d. for* $p_{m,n}$ *given*

$$T_{m-1}(p_{m,m-1}) = t_{m-1}, \dots, T_{n+1}(p_{m,n+1}) = t_{n+1}, T_n(p_{m,n}) = t_n$$

is Q_{n,t_n}.

Proof: The plan is to fix n and do an induction on m. The case $m = n+1$ is just assumption (4.2). Suppose the result for some $m \geqslant n+1$. To proceed inductively, it is necessary to compute, relative to $Q_{m+1,t}$, an r.c.d. for $p_{m+1,n}$ given

$$T_m(p_{m+1,m}) = t_m, T_{m-1}(p_{m+1,m-1}) = t_{m-1}, \dots, T_n(p_{m+1,n}) = t_n. \quad \dots \quad (4.4)$$

The idea is to use Lemma 4.3, and condition on $T_m(p_{m+1,m})$ first. Put p_{m+1} for X, so $\Omega_X = \mathscr{X}_m$. Also put $p_{m,n}$ for f, and T_m for g. For h, put the vector

$$T_{m-1}(p_{m,m-1}), \dots, T_n(p_{m,n}).$$

The composition $h(X)$ is easily computed because

$$p_{m+1,m} = p_{m+1}, p_{m,m-1} \circ p_{m+1} = p_{m+1,m-1}, \dots, p_{m,n} \circ p_{m+1} = p_{m+1,n}. \quad \dots \quad (4.5)$$

Relative to $Q_{m+1,t}$, an r.c.d. for X given $g(X) = t_m$ is just Q_{m,t_m}, by assumption (4.2). The next step is to compute, relative to Q_{m,t_m}, an r.c.d. for f given h. By the inductive assumption, however, this is just Q_{n,t_n}, which is certainly a measurable function of $(t_n, t_{n+1}, \dots, t_m)$. By Lemma 4.3 and the identity (4.5), relative to $Q_{m+1,t}$, an r.c.d. for $p_{m+1,n}$ given the equalities (4.4) is Q_{n,t_n}. □

This lemma must now be translated into a statement about the partially exchangeable probability P on $(\mathscr{X}, \mathscr{A})$.

Lemma 4.5 : *Let $P \in M_Q$ and $m > n$. Relative to P, an r.c.d. for ξ_n given*

$$T_n \circ \xi_n = t_n, \ldots, T_m \circ \xi_m = t_m$$

is just Q_{n, t_n}.

Proof: Again, Lemma 4.3 can be used. Put ξ_m for X, $p_{m,n}$ for f, T_m for g, and for h the vector

$$T_n \circ p_{m,n}, \ldots, T_{m-1} \circ p_{m, m-1}.$$

Of course, $f(X) = p_{m,n} \circ \xi_m = \xi_n$. The requisite r.c.d. for $f(X)$ can be computed by conditioning first on $g(X)$, then on $h(X)$. By the definition of M_Q, the first conditioning gives Q_{m, t_m}. The second gives Q_{n, t_n} by Lemma 4.4. □

An immediate consequence is the following :

Lemma 4.6 : *Let $P \in M_Q$. Relative to P, a r.c.d. for ξ_n given $\Sigma^{(n)}$ is $Q_{n, T_n(\xi_n)}$.*

The next lemma establishes the convergence of Q_{n, t_n} along almost all subsequences.

Lemma 4.7 : *Let $\mathscr{X}_L$ be the set of $x \in \mathscr{X}$ such that for each k, as $n \to \infty$, $Q_{n, T_n[\xi_n(x)]}\, p_{n,k}^{-1}$ converges weak-star to a limiting probability on $(\mathscr{X}_k, \mathscr{A}_k)$; denote this limit by $Q(k, x, \cdot)$. Then $\mathscr{X}_L \in \Sigma$ and $P(\mathscr{X}_L) = 1$ for all $P \in M_Q$. Furthermore, $x \to Q(k, x, A)$ is Σ-measurable for each $A \in \mathscr{A}_k$, and a version of $P\{\xi \in A \mid \Sigma\}$ for any $P \in M_Q$.*

As a matter of notation, $Q(k, x, \cdot)$ is to be distinguished from $Q_{k, t}$. The latter is the distribution of ξ_k given $T_k(\xi_k)$. The former is the distribution of ξ_k given the tail σ-field of $\{T_n(\xi_n)\}$. Note that $Q(k, x, \cdot)$ does not depend on P; this is important.

Proof: It is at this point that the Polishness of $\mathscr{X}_k$ is used. Embed $\mathscr{X}_k$ as a Borel subset of the compact metric space $\hat{\mathscr{X}}_k$. Let $C(\hat{\mathscr{X}}_k)$ denote the space of continuous functions on $\hat{\mathscr{X}}_k$; if f is a function on $\hat{\mathscr{X}}_k$, its restriction to $\mathscr{X}_k$ will still be written as f. Let $C_0(\hat{\mathscr{X}}_k)$ be a countable dense subset of $C(\hat{\mathscr{X}}_k)$, in the sup norm.

Let x denote a typical point in $\mathscr{X}$, and y_k a typical point in $\mathscr{X}_k$. For any bounded measurable function f on $\mathscr{X}_k$,

$$x \to \int_{\mathscr{X}_k} f(y_k) Q_{n, T_n[\xi_n(x)]}\, p_{n,k}^{-1}(dy_k) \qquad \ldots \quad (4.6)$$

is a version of $\qquad E_P\{f(\xi_k) \mid T_n(\xi_n), T_{n+1}(\xi_{n+1}), \ldots\}$

by Lemma 4.6; indeed, $\xi_k = p_{n,k}(\xi_n)$. The function (4.6) must converge P-almost surely to $E_P\{f(\xi_k) \mid \Sigma\}$, by the backwards martingale convergence theorem. This will now be restated as follows. Let G_{kf} be the set of x's where the function in (4.6) converges to a limit as $n \to \infty$; call the limit $\lambda(k, x, f)$. Then $G_{kf} \in \Sigma$ and $P(G_{kf}) = 1$. Also, $\lambda(k, \cdot, f)$ is Σ-measurable, and a version of $E_P\{f(\xi_k) \mid \Sigma\}$. Note that G_{kf} and $\lambda(k, x, f)$ do not depend on P.

Let G_k be the intersection of G_{kf} over $f \in C_0(\hat{\mathscr{X}}_k)$. Again $G_k \in \Sigma$; and $P(G_k) = 1$ for all $P \in M_Q$. Fix $x \in G_k$, abbreviate μ_n for the probability $Q_{T_n[\xi_n(x)]}\, p_{n,k}^{-1}$ on $(\mathscr{X}_k, \mathscr{F}_k)$. Consider now μ_n as a probability on $\hat{\mathscr{X}}_k$. Then the sequence $\{\mu_n\}$ is pre-compact in the weak-star topology, and $\int f d\mu_n \to \lambda(k, x, f)$ for every $f \in C_0(\hat{\mathscr{X}}_k)$. The conclusion is that μ_n converges weak-star to $Q(k, x, \cdot)$, a probability on $\hat{\mathscr{X}}_k$; furthermore,

$$x \to \int_{\hat{\mathscr{X}}_k} f(y_k) Q(k, x, dy_k) \qquad \ldots \quad (4.7)$$

is Σ-measurable, and a version of $E_P\{f(\xi_k) \mid \Sigma\}$ for all $f \in C_0(\hat{\mathscr{X}}_k)$.

Does $Q(k, x, \mathscr{X}_k) = 1$? By a standard argument, for any bounded Borel function on $\hat{\mathscr{X}}_k$, the function in (4.7) is Σ-measurable, and for any $S \in \Sigma$,

$$\int_{G_k \cap S} \int_{\hat{\mathscr{X}}_k} f(y_k) Q(k, x, dy_k) P(dx) = \int_{\mathscr{X} \cap S} f(\xi_k) dP. \qquad \ldots \quad (4.8)$$

Putting $f = 1_{\mathscr{X}_k}$ and $S = \mathscr{X}$ proves that

$$H_k = \{x : x \in G_k \text{ and } Q(k, x, \mathscr{X}_k) = 1\} \in \Sigma$$

has P-measure 1 for all $P \in M_Q$. Now

$$\mathscr{X}_L = \bigcap_k H_k.$$

For $x \in \mathscr{X}_L$, view $Q(k, x, \cdot)$ as a probability on $(\mathscr{X}_k, \mathscr{F}_k)$. As a function of x, this is Σ-measurable, and an r.c.d. for ξ_k given Σ by (4.8). □

Lemma 4.8 : *Let $\mathscr{X}_C$ be the set of $x \in \mathscr{X}_L$ such that for all k,*

$$Q(k, x, \cdot) p_k^{-1} = Q(k-1, x, \cdot).$$

Then $\mathscr{X}_C \in \Sigma$ and $P(\mathscr{X}_C) = 1$ for all $P \in M_Q$.

Proof: That $\mathscr{X}_C \in \Sigma$ is clear. For $f \in C_0(\hat{\mathscr{X}}_{k-1})$, as defined in the proof of the previous lemma,

$$x \to \int_{\hat{\mathscr{X}}_{k-1}} f(y_{k-1}) Q(k-1, x, dy_{k-1}) \qquad \ldots \quad (4.8a)$$

is a version of $E_P\{f(\xi_{k-1})\,|\,\Sigma\}$. Now $f(p_k)$ is a bounded measurable function on $\mathscr{X}_k$. So

$$x \to \int_{\mathscr{X}_k} f[p_k(y_k)]Q(k, x, dy_k) \tag{4.8b}$$

is a version of $E_P\{f[p_k(\xi_k)]\,|\,\Sigma\} = E_P\{f(\xi_{k-1})\,|\,\Sigma\}$, because $p_k(\xi_k) = \xi_{k-1}$. Let H_{kf} be the set of x's for which (4.8a) and (4.8b) agree. Then $H_{kf} \in \Sigma$ and $P(H_{kf}) = 1$. Now

$$\mathscr{X}_C = \bigcap_{k,f} H_{kf},$$

where f runs over $C_0(\hat{\mathscr{X}}_k)$. □

For each $x \in \mathscr{X}_C$, the probabilities $Q(k, x, \cdot)$ are consistent as k varies. So there is a unique countably additive probability $Q(x, \cdot)$ on $(\mathscr{X}, \mathscr{A})$ such that for any k and $A \in \mathscr{A}_k$,

$$Q(x, \{\xi_k \in A\}) = Q(k, x, A).$$

Again, the Polishness of the $\mathscr{X}_k$ is used.

Lemma 4.9 : *Fix $A \in \mathscr{A}$. The function $x \to Q(x, A)$ on $\mathscr{X}_C$ is Σ-measurable, and a version of $P(A\,|\,\Sigma)$ for any $P \in M_Q$.*

Lemma 4.10 : *Let $\mathscr{X}_Q$ be the set of $x \in \mathscr{X}_C$ such that $Q(x, \cdot) \in M_Q$. Then $\mathscr{X}_Q \in \Sigma$ and $P(\mathscr{X}_Q) = 1$ for all $P \in M_Q$.*

Proof : Recall that T_n is a measurable map from $\mathscr{X}_n$ into W_n, where $\mathscr{X}_n$ is equipped with the Borel σ-field $\mathscr{A}_n$, and W_n with the Borel σ-field $\mathscr{B}_n$. Write y for a typical point in $\mathscr{X}$. Now x in $\mathscr{X}_C$ is also in $\mathscr{X}_Q$ iff

$$\int_{T_n[\xi_n(y)] \in B} Q_{n,T_n[\xi_n(y)]}(A)Q(x, dy) = Q(x, \{T_n(\xi_n) \in B \text{ and } \xi_n \in A\}) \quad \ldots \quad (4.9)$$

for all n, all $B \in \mathscr{B}_n$ and $A \in \mathscr{A}_n$. Here, A and B can be restricted to countable generating algebras for their respective σ-fields; and both sides of (4.9) are Σ-measurable functions of x : so $\mathscr{X}_Q \in \Sigma$.

To show that $P(\mathscr{X}_Q) = 1$, it is enough to show that both sides of (4.9) have the same P-integral over arbitrary sets $G \in \Sigma$. Now the P-integral of the right side over G is

$$P\{G \text{ and } T_n(\xi_n) \in B \text{ and } \xi_n \in A\},$$

by virtue of Lemma 4.9, with $\{T_n(\xi_n) \in B \text{ and } \xi_n \in A\}$ for A. Likewise, the P-integral of the left side over G is

$$E_P\{1_G \cdot 1_B[T_n(\xi_n)] \cdot Q_{n,T_n(\xi_n)}(A)\}.$$

But $G \in \Sigma^{(n)}$, as is $\{T_n(\xi_n) \in B\}$, so the last display equals

$$P\{G \text{ and } T_n(\xi_n) \in B \text{ and } \xi_n \in A\},$$

by Lemma 4.6. □

We have now reached the point where Dynkin's theorem (1978) can be applied. For ease of reference, the argument will be sketched.

Lemma 4.11 : *Let* Σ_Q *be the* σ*-field of subsets of* $\mathscr{X}_Q$ *generated by* $x \to Q(x, A)$ *as* A *ranges over* $\mathscr{A}$. *Then* Σ_Q *is countably generated.*

Proof : A countable generating class of sets is $\{x : x \in \mathscr{X}_Q \text{ and } Q(x, A) > r\}$ where A runs over a countable field generating $\mathscr{A}$ and r runs over the rationals. □

Lemma 4.12 : *For* $A \in \Sigma$, *let* A^* *be the set of* $x \in \mathscr{X}_Q$ *with* $Q(x, A) = 1$. *Then* $A^* \in \Sigma_Q$, *and* $P(A \Delta A^*) = 0$ *for all* $P \in M_Q$.

Proof : The function $x \to Q(x, A)$ is a version of $P(A \mid \Sigma)$ by Lemma 4.9, and $P(A \mid \Sigma) = 1_A$ with P-probability 1. □

The inseparable σ-field Σ is therefore the same (up to null sets) as the separable sub-σ-field Σ_Q. In some applications, it is important that A^* does not depend on P.

Lemma 4.13 : *For* $x \in \mathscr{X}_Q$, *let* $\Sigma_Q(x)$ *be the* Σ_Q*-atom containing* x, *namely*

$$\{y : y \in \mathscr{X}_Q \text{ and } Q(y, \cdot) = Q(x, \cdot)\}.$$

Let $\mathscr{X}_E$ *be the set of* $x \in \mathscr{X}_Q$ *such that* $Q(x, \Sigma_Q(x)) = 1$. *Then* $\mathscr{X}_E \in \Sigma_Q$ *and* $P(\mathscr{X}_E) = 1$ *for all* $P \in M_Q$.

Proof : Let A run through a countable field generating Σ_Q. Then $\mathscr{X}_E$ is

$$\bigcap_A \{x : x \in \mathscr{X}_Q \text{ and } Q(x, A) = 1_A(x)\}.$$

The proof is completed by appealing to Lemma 4.9, as before. □

It may be useful to observe that for $x \in \mathscr{X}_Q$,

$$x \in \mathscr{X}_E \text{ iff } Q(x, A) = 1_A(x) \text{ for all } A \in \Sigma_Q. \quad \ldots \quad (4.10)$$

In general, the $Q(x, \cdot)$-measure of the Σ-atom through x is 0, for the measure may be continuous and the atom is countable. By contrast, the Σ_Q-atom will usually be much larger, having the power of the continuum.

Lemma 4.14 : *If* $x \in \mathscr{X}_E$, *then* $Q(x, \cdot)$ *is* $0-1$ *on* Σ.

Proof : If $x \in \mathscr{X}_E$, then $Q(x, \cdot)$ is $0-1$ on Σ_Q. But $Q(x, \cdot) \in M_Q$; see Lemma 4.10. So $Q(x, \cdot)$ is $0-1$ on Σ by Lemma 4.12. □

Lemma 4.15 : *Let* $\Sigma_E = \mathscr{X}_E \bigcap \Sigma_Q$, *viz., the* σ*-field of subsets of* $\mathscr{X}_E$ *of the form* $\mathscr{X}_E \bigcap A$ *with* $A \in \Sigma_Q$. *For* $P \in M_Q$, *there is one and only one probability measure* $\hat{P}$ *on* $(\mathscr{X}_E, \Sigma_E)$ *such that*

$$P = \int_{\mathscr{X}_E} Q(x, \cdot)\hat{P}(dx); \quad \ldots \quad (4.11)$$

namely, $\hat{P}$ is the restriction of P to Σ_E. In particular, if P and P' in M_Q agree on Σ_E, then $P = P'$.

Proof: If $\hat{P}$ is the restriction of P to Σ_E, then (4.11) holds by Lemma 4.9. Conversely, suppose (4.11) holds for some probability $\hat{P}$ on $(\mathcal{X}_E, \Sigma_E)$. Then for $A \in \Sigma_E$,

$$P(A) = \int_{\mathcal{X}_E} Q(x, A)\hat{P}(dx) \qquad \hat{P}(A) = \int_{\mathcal{X}_E} 1_A(x)\hat{P}(dx)$$

$$= \int_{\mathcal{X}_E} 1_A(x)\hat{P}(dx) \qquad = \int_{\mathcal{X}_E} Q(x, A)\hat{P}(dx) \text{ by (4.10)}$$

$$= \hat{P}(A), \qquad = P(A) \qquad \text{by (4.11).} \ \square$$

Remark: Suppose $\mathcal{X}_n$ is compact and non-empty, p_n and T_n continuous, and $Q_{n,t}$ is a weak-star continuous function of t. Then M_Q is a non-empty compact convex set, and Choquet's theorem applies, see Lemma 4.16 below for the identification of the extreme points. Under these conditions, $\mathcal{X}_L = \mathcal{X}_Q = \mathcal{X}_C$. Apparently, $\mathcal{X}_E$ may be smaller than $\mathcal{X}_C$, but we do not have an example. Without compactness and continuity conditions, M_Q can be empty : see example 3.4.

Lemma 4.16. *The probability $P \in M_Q$ is extreme iff it is $0-1$ on Σ; alternatively, P is extreme iff*

$$P\{x : x \in \mathcal{X}_L \text{ and } Q(x, \cdot) = P\} = 1. \qquad \dots \quad (4.12)$$

In particular, as x runs over $\mathcal{X}_E$, the probability $Q(x, \cdot)$ runs over the extreme P in M_Q.

Proof: First, P is $0-1$ on Σ iff it is $0-1$ on Σ_E, by Lemmas 4.11–12. Next, P is $0-1$ on Σ_E iff it concentrates on an atom, namely, a set of the form

$$\{x : x \in \mathcal{X}_E \text{ and } Q(x, \cdot) = P'\};$$

then $P' = P$ by (4.11) : that is, P is $0-1$ on Σ_E iff (4.12) holds.

If P is extreme, then P must be $0-1$ on Σ_E; if not, let $A \in \Sigma_E$ with $0 < P(A) < 1$. Then

$$P = \alpha P_1 + (1-\alpha)P_2, \qquad \dots \quad (4.13)$$

where

$$\alpha = P(A),\ P_1 = \alpha^{-1} \int_A Q(x, \cdot)\hat{P}(dx),\ P_2 = (1-\alpha)^{-1} \int_{\mathcal{X}_E - A} Q(x, \cdot)\hat{P}(dx).$$

Of course, $P_1(A) = 1$ and $P_2(A) = 0$ by (4.10); so $P_1 \neq P_2$ and P is not extreme. This contradiction proves that P is $0-1$ on Σ_E.

Conversely, if P is 0—1 on Σ_E, then P is extreme; if not, (4.13) holds for some α with $0 < \alpha < 1$ and some pair $P_1 \neq P_2$ in M_Q Clearly, $P_2 << P$ on Σ_E. So, P_1 and P_2 and P all concentrate on the same Σ_E-atom, i.e., $P_1 = P_2 = P$ on Σ_E, and hence on $\mathscr{A}$ by Lemma 4.15. This contradiction shows that P is extreme.

If $x \in \mathscr{X}_E$, then $Q(x, \cdot)$ is extreme, by (4.12) and the definition of $\mathscr{X}_E$ in Lemma 4.13. Conversely, if P is extreme, then P is 0—1 on Σ_E and $P\{x : Q(x, \cdot) = P\} = 1$ by (4.12). □

This completes the proof of Theorem 4.1. Now Theorem 1.1 can be derived as a special case. Let $\mathscr{X}_n = \prod_{i=1}^{n} \Omega_i$ and $\mathscr{A}_n = \prod_{i=1}^{n} \mathscr{F}_i$, so $\mathscr{X} = \prod_{i=1}^{\infty} \Omega_i$ and $\mathscr{A} = \prod_{i=1}^{\infty} \mathscr{F}_i$ and $\xi_n = (X_1, \ldots, X_n)$. With these identificatoins, the two definitions of M_Q coincide. Theorem 1.1 involves the σ-field $\hat{\Sigma} = \bigcap_{n=1}^{\infty} \hat{\Sigma}^{(n)}$,

where
$$\hat{\Sigma}^{(n)} = \sigma\{T_n(X_1, \ldots, X_n), X_{n+1}, \ldots\}.$$

Theorem 4.1 involves the σ-field $\Sigma = \bigcap_{n=1}^{\infty} \Sigma^{(n)}$, where

$$\Sigma^{(n)} = \sigma\{T_n(X_1, \ldots, X_n), T_{n+1}(X_1, \ldots, X_{n+1}), \ldots\}.$$

It will be shown that Σ is smaller than $\hat{\Sigma}$, but equivalent up to null sets for any $P \in M_Q$. This will prove Theorem 1.1, with $\mathscr{X}_E$ for E.

Lemma 4.17 : *Assume condition* (1.1–3). *Then* $\Sigma^{(n)} \subset \hat{\Sigma}^{(n)}$ *for all* n, *and* $\Sigma \subset \hat{\Sigma}$.

Proof : $\Sigma^{(n)} \subset \hat{\Sigma}^{(n)}$ for all n, because e.g., $T_{n+1}(X_1, \ldots, X_{n+1})$ can be computed from $T_n(X_1, \ldots, X_n)$ and X_{n+1} by condition (1.2). To make this rigorous, use Blackwell's (1954) theorem on saturation. This will show that for any $m > n$,

$$\Sigma^{(n,m)} \subset \hat{\Sigma}^{(n,m)},$$

where

$$\Sigma^{(n,m)} = \sigma\{T_n(X_1, \ldots, X_n), T_{n+1}(X_1, \ldots, X_{n+1}), \ldots, T_m(X_1, \ldots, X_m)\},$$
$$\hat{\Sigma}^{(n,m)} = \sigma\{T_n(X_1, \ldots, X_n), X_{n+1}, \ldots, X_m\}. \quad \ldots \quad (4.14)$$

Indeed, both σ-fields are Borel, and the atoms of $\hat{\Sigma}^{(n,m)}$ are smaller. □

Lemma 4.18 : *Assume condition* (1.1–3). *Fix* $P \in M_Q$. *Then* $Q_{n,T_n(X_1,\ldots,X_n)}$ *is an r.c.d. for* $(X_1, \ldots, X_n)$ *given* $\hat{\Sigma}^{(n)}$.

Proof : It will be shown first that $Q_{n, T_n(X_1, \ldots, X_n)}$ is an r.c.d. for $(X_1, \ldots, X_n)$ given $\hat{\Sigma}^{(n,m)}$, as defined in (4.14); then m can be sent to ∞. The argument will only be sketched, it is similar to the one in Lemmas 4.4–5. First, the σ-field $\hat{\Sigma}^{(n,m)}$ coincides with the apparently larger σ-field spanned by

$$T_n(X_1, \ldots, X_n), T_{n+1}(X_1, \ldots, X_{n+1}), \ldots, T_m(X_1, \ldots, X_m)\ X_{n+1}, \ldots, X_m.$$

An r.c.d. for $(X_1, \ldots, X_n)$ given this menu of variables can be computed by using Lemma 4.3, conditioning first that $T_m(X_1, \ldots, X_m) = t_m$; the definition of M_Q is used to accomplish this first conditioning. The problem is reduced to computing, relative Q_{m, t_m}, an r.c.d. for $(X_1, \ldots, X_n)$ given

$$T_n(X_1, \ldots, X_n), T_{n+1}(X_1, \ldots, X_{n+1}), \ldots, T_{m-1}(X_1, \ldots, X_{m-1})\ X_{n+1}, \ldots, X_{m-1}, X_m.$$

Lemma 4.3 can be used again : condition first on $T_{m-1}(X_1, \ldots, X_{m-1}) = t_{m-1}$ and X_m, using assumption (1.3). The problem is reduced to computing, relative to $Q_{m-1, t_{m-1}}$, an r.c.d. for $(X_1, \ldots, X_n)$ given

$$T_n(X_1, \ldots, X_n), T_{n+1}(X_1, \ldots, X_{n+1}), \ldots, T_{m-2}(X_1, \ldots, X_{m-2})$$
$$X_{n+1}, \ldots, X_{m-2}, X_{m-1}.$$

Proceeding in this way, one verifies the claim. □

Proposition 4.1 : *Assume conditions* (1.1–3). *For* $A \in \hat{\Sigma}$, *let* $\tilde{A}$ *be the set of* $x \in \mathcal{X}_E$ *such that* $Q(x, A) = 1$. *Then* $\tilde{A} \in \Sigma$, *and* $P(A \Delta \tilde{A}) = 0$ *for all* $P \in M_Q$.

Proof : As noted in Lemma 4.17, $\Sigma^{(n)} \subset \hat{\Sigma}^{(n)}$. And since $Q_{n, T_n(X_1, \cdots, X_n)}$ is $\Sigma^{(n)}$-measurable, it is an r.c.d. for $(X_1 \ldots X_n)$ given $\Sigma^{(n)}$, as well as an r.c.d. for $(X_1 \ldots X_n)$ given $\hat{\Sigma}^{(n)}$: see Lemma 4.18 or Lemma 4.6.

Turn now to Lemma 4.7 : one verifies that for $x \in \mathcal{X}_L$: $Q(k, x, \cdot)$ is an r.c.d. for $(X_1, \ldots, X_k)$ either given Σ or given $\hat{\Sigma}$. Likewise, in Lemma 4.9, $Q(x, \cdot)$ is an r.c.d. either given Σ or given $\hat{\Sigma}$. For $A \in \hat{\Sigma}$, then $Q(x, A) = 1_A(x)$ a.e. □

Technical note : In connection with example 2.1, it is pointed out that $\mathcal{X}_E = \mathcal{X}_Q = \mathcal{X}_C = \mathcal{X}_L$, and Σ_E is spanned by Λ. Also,

$$\hat{\Sigma}^{(n)} = \sigma\{T_n(X_1, \ldots, X_n), X_{n+1}, \ldots\}$$

while $$\hat{\Sigma}^{(n)} = \sigma\{T_n(X_n, \ldots, X_n), T_{n+1}(X_1, \ldots, X_{n+1}), \ldots\}.$$

A moment's thought shows that $T_{n+1}(X_1, ..., X_{n+1})$ and $T_n(X_1, ..., X_n)$ determine X_{n+1}, so in fact $\Sigma^{(n)} = \hat{\Sigma}^{(n)}$: this is the σ-field of measurable sets invariant under permutations of the first n-coordinates. In particular, $\hat{\Sigma} = \Sigma$ is the σ-field of exchangeable events.

References

The early history of exchangeability is covered in Hewitt and Savage (1955). Kingman (1978) brings this up to date. We have collected other references in the bibliography to this paper.

ALDOUS, D. (1977) : Limit theorems for subsequences of arbitrarily dependent sequences of random variables. *Z. Wahr. verw. Geb.* 40, 59-82.

ALDOUS, D. (1981a) : Partial exchangeability and $\bar{d}$-topologies. In *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds) North Holland, Amsterdam 23–38.

——— (1981b) : Representations for partially exchangeable arrays of random variables. *Jour. Mult. Anal.* 11, 581-598.

ALDOUS D. (1984) : Exchangeability and related topics. *In Ecole d'ete st. Fleur* 1983, *ed. Henneque.* To appear, Springer lecture Notes in Mathematics.

ALDOUS, D. and PITMAN, J. W. (1979) : On the zero-one law for exchangeable events. *Ann. Prob.* 7, 704-723.

ANDERSEN, E. B. (1970) : Sufficiency and exponential families for discrete sample spaces. *Jour. Amer. Statist. Assoc.* 65, 1248-1255.

BAHADUR, R. R. (1954) : Sufficiency and statistical decision functions. *Ann. Math. Statist.* 25, 423-462.

BERG, C. and RESSEL, P. (1978) : Une forme abstraite du theoreme de Schoenberg. *Arch. Math.* 30, 55-61.

BERMAN, S. (1962a) : Equally correlated random variables. *Sankhyā A* 24, 155-156.

——— (1962b) : An extension of the arc sine law. *Ann. Math. Statist.* 33, 681-684.

——— (1962c) : Limiting distributions of the maximum term in sequences of dependent random variables. *Ann. Math. Statist.* 33, 894-908.

——— (1965) : Sign-invariant random variables and stochastic processes with sign-invariant increments. *Trans. Amer. Math. Soc.* 119,216-243.

——— (1967) : Sign-invariant random elements in Hilbert space. *Z. Wahr. verw. Geb.* 8, 73-82.

——— (1980) : Stationarity, isotropy, and sphericity in l_p. *Z. Wahr. verw. Geb.* 54, 21-23.

BLACKWELL, D. (1954) : On a class of probability spaces. *Proc. 3rd. Berkeley Symp.* Vol. 2, 1-6.

BLUM, J. R., CHERNOFF, H. R., ROSENBLATT, M. and TEICHER, H. (1958) : Central limit theorems for interchangeable processes. *Canad. Jour. Math.* 10, 222-229.

BRETAGNOLLE, Y., DACUNHA-CASTELLE and KRIVINE, Y. L. (1966) : Lois stables et espaces L^p. *Ann. Inst. H. Poincare, Sect B II*, 3, 231-259.

CHOQUET, G. (1969) : *Lectures on Analysis*, Vol. iii. Benjamin : Reading, Massachusetts.

DAWID, A. P. (1977) : Extendibility of spherical matrix distributions. *J. Multivar. Anal.* 8, 567-572.

DE FINETTI, B. (1931) : Funzione Caratteristica di un fenomeno aleatorio. *Atti della R. Accademia Nazionale dei Lincii Ser. 6, Memorie, classe di Scienze, Fisiche, Matematiche e Naturali*, Vol. 4, 251-299.

——— (1937) La Prévision : Ses lois logiques ses sources subjectives, *Ann. de l'institut Henri Poincaré* 7, 1-68. Translated in *Studies in Subjective Probability*, H. Kyburg and H. Smokler, Eds., Wiley : New York (1964).

——— (1938) : Sur la condition d'equivalence partielle. *Actualités Scientifiques et Industrielles*, No. 739. Herman and Cii : Paris. Translated in *Studies in Inductive Logic and Probability, II*, R. Jeffrey, Ed. University of California Press : Berkeley.

——— (1972) : *Probability, Induction, and Statistics.* Wiley : New York.

DELLACHERIE, C. and MEYER, P. A. (1975) : *Probabilities et Potentiel.* Herman : Paris.

DIACONIS, P. (1977) : Finite forms of de Finetti's theorem on exchangeability. *Synthese*, 36, 271-281.

——— (1980) : Average running time of the fast Fourier transform. *J. Algorithms* 1, 187-208.

DIACONIS, P. and FREEDMAN, D. (1980a) : de Finetti's theorem for Markov chains. *Ann. Prob.* 8, 115-130.

——— (1980b) : de Finetti's generalizations of exchangeability. In *Studies in Inductive Logic and Probability* Vol. II. R. C. Jeffrey, ed., University of California Press, Berkeley.

——— (1980c) : Finite exchangeable sequences. *Ann. Prob.* 8, 745-764.

——— (1981) : On the statistics of vision : the Julesz conjecture. *Jour. Math. Psychol.* 24, 112-138.

——— (1983) : Finite forms of some versions of de Finetti's theorem. Unpublished manuscript

DUBINS, L. and FREEDMAN, D. (1979) : Exchangeable processes need not be mixtures of independent identically distributed random variables. *Z. Wahr. verw. Geb.* 48, 115-132.

DYNKIN, E. (1978) : Sufficient statistics and extreme points. *Ann. Prob.* 6, 705-730.

EATON, M. (1981) : On the projections of isotropic distributions. *Ann. Statist.* 9, 391-400.

EVERITT, C. W. F. (1974) : *James Clerk Maxwell.* Scribner's : New York.

FARRELL, R. H. (1962) : Representation of invariant measures. *Ill. Jour. Math.* 6, 447-467.

FOLMER, H. (1975) : Phase transition and Martin boundary. *Springer Lecture Notes in Mathematics* 465.

FREEDMAN, D. (1962a) : Mixtures of Markov processes. *Ann. Math. Statist.* 33, 114-118.

——— (1962b) : Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Statist.* 33, 916-923.

——— (1963) : Invariants under mixing which generalize de Finetti's theorem : Continuous time parameter. *Ann. Math. Statist.* 34, 1194-1216.

——— (1980) : A mixture of independent identically distributed random variables need not admit a regular conditional probability given the exchangeable σ-field. *Z. Wahr. verw. Geb.* 51, 239-248.

FREEDMAN, D., and DIACONIS, P. (1982) : de Finetti's theorem for symmetric location families. *Ann. Statist.* 10, 184-189.

GEORGII, H. O. (1975) : Canonical Gibbs states, their relation to Gibbs states, and applications to two-valued Markov chains. *Z. Wahr. verw. Geb.* 32, 272-300.

——— (1979) : Canonical Gibbs Measures. *Springer Lecture Notes in Mathematics* 760.

HEATH, D. and SUDDERTH, W. (1976) : de Finetti's theorem for exchangeable random variables. *Amer. Statistician* 30, 188-189.

HEWITT, E. and SAVAGE, L. J. (1955) : Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* 80, 470-501.

HINKLEY, D. V. (1976): Predictive likelihood. *Ann. Statist.* 7, 718-728.

HIPP, C. (1974): Sufficient statistics and exponential families. *Ann. Statist.* 2, 1283-1292.

HOOVER, D. (1981a): Relations on probability spaces, and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, New Jersey.

——— (1981b): Pairwise Independence Implies Independence almost everywhere. Preprint. Institute for Advanced Study, Princeton, New Jersey.

HUNT, G. (1960): Markoff chains and Martin boundaries. *Ill, Jour. Math.* 4, 313-340.

JAYNES, E. (1982): Some applications and extensions of the de Finetti representation theorem. To appear in *Bayesian Inference and Decision Techniques with Applications:* Essays in Honor of Bruno de Finetti. Prem K. Goel and Arnold Zellner, Editors. North-Holland: Amsterdam.

KALLENBERG, O. (1973): Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahr. verw. Geb.* 27, 23-26.

——— (1974): Path properties of processes with independent and interchangeable increments. *Z. Wahr. verw. Geb.* 28, 257-271.

——— (1975): On symmetrically distributed random measures. *Trans. Amer. Math. Soc.* 202, 105-121.

——— (1975): Infinitely divisible processes with interchangeable increments and random measures under convolution. *Z. Wahr. verw. Geb.* 32, 309-321.

——— (1976): *Random Measures.* Academic Press: Berlin.

——— (1981): Characterizations and Embedding Properties in Exchangeability. Technical Report No. 10, Department of Mathematics, Chalmers University of Technology, Göteborg, Sweden.

KENDALL, D. (1963): Extreme point methods in stochastic analysis. *Z. Wahr. verw. Geb.* 1, 295-300.

——— (1967): On finite and infinite sequences of exchangeable events. *Studia Sci. Math. Hungar,* 2.

KERSTAN, J., and WAKOLBINGER, A. (1981): Ergodic decomposition of probability. *Z. Wahr. verw. Geb.* 56, 399-414.

KINGMAN, J. (1978): Uses of exchangeability. *Ann. Prob.* 6, 183-197.

LANFORD, O. and RUELLE, D. (1969): Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.* 13, 194-215.

LAURITZEN, S. L. (1972): Sufficiency and time series analysis. *University of Copenhagen, Institute of Mathematical Statistics Preprint No.* 11.

——— S. L. (1974a): On the interrelationships among sufficiency and some related concepts. *University of Copenhagen Institute of Mathematical Statistics Preprint No.* 8

——— (1974b): Sufficiency, prediction and extreme models. *Scand. J. Statist.* 1, 128-134.

——— (1975): General exponential models for discrete observations. *Scand. J. Statist.* 2, 23-33.

——— (1976): Projective statistical fields. University of Copenhagen, Institute of Mathematical Statistics Preprint No. 6.

——— (1982): *Statistical Models as Extremal Families.* Aalborg University Press, Aalborg.

LEHMANN, E. (1959): *Testing Statistical Hypotheses.* Wiley: New York.

LETAC, G. (1981): Isotropy and sphericity: Some characterizations of the normal distribution. *Ann. Statist.* 9, 408-417.

MAITRA, A. (1977): Integral representations of invariant measures. *Trans. Amer. Math. Soc.* 229, 209-225.

MANDELBROT, B. (1962): The role of sufficiency and of estimation in thermodynamics. *Ann. Math. Statist.* 33, 1021-1038.

MARTIN-LÖF, P. (1970): Statistika Modeller. Notes by Rolf Sundberg. Mimeographed lecture notes.

——— (1974): Repetitive structures and the relation between canonical and micro-canonical distributions in statistics and statistical mechanics. In *Proceedings of Conference of Foundational Questions in Statistical Inference,* O. Barndorff-Nielsen, P. Blaesild, G. Schon, Editors, Aarhus.

——— (1975): Reply to Sverdrup's polemical article tests without power. *Scand. J. Statist.* 2, 161-165.

OLSHEN, R. A. (1971): The coincidence of measure algebras under an exchangeable probability. *Z. Wahr. verw. Geb.* 18, 153-158.

——— (1974): A note on exchangeable sequences. *Z. Wahr. verw. Geb.* 28, 317-324.

OXTOBY, J. C. (1952): Ergodic sets. *Bull. Amer. Math. Soc.* 58, 116-136.

PARTHASARATHY, K. R. (1973): On the uniqueness of symmetric quantum processes. *Jour. Indian Math. Soc.* 37, 35-42.

PHELPS, R. R. (1966): *Lectures on Choquet's Theorem.* Van Nostrand: Princeton, New Jersey.

PRESTON, C. (1976): Random Fields. *Springer Lecture Notes in Mathematics* 534.

——— (1979): Canonical and microcanonical Gibbs states. *Z. Wahr. verw. Geb.* 46, 125-158.

PITMAN, J. (1978): An extension of de Finetti's theorem. *Advances in Appl. Prob.* 10, 268-270.

RESSEL, PAUL. (1983): DE FINETTI-TYPE THEOREMS: an analytical approach. *Technical Report* Department of Statistics, Stanford University.

RUELLE, D. (1978): *Thermodynamic Formalism.* Addison-Wesley: Reading, Massachusetts.

SCHOENBERG, I. J. (1938): Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.* 44, 522-536.

SMITH, A. M. F. (1981): On random sequences with centered spherical symmetry. *Jour. Roy. Statist. Soc. Ser. B,* 208-209.

SOKAL, A. (1981): Existence of compatible families of proper regular conditional probabilities. *Z. Wahr. verw. Geb.* 56, 537-548.

VARADARAJAN, V. S. (1963): Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* 109, 191-220.

ZABELL, S. (1982): W. E. Johnson's "Sufficientness" postulate. *Ann. Statist.* 10, 1091-1099.

ZAMAN, A. (1981): An Approximation Theorem for Finite Markov Exchangeability. Ph.D. thesis, Department of Statistics, Stanford University.

ZAMAN, A. (1984): Urn models for Markov exchangeability. *Ann. Prob.* 12, 223-229.